Optimization of Adaptive Enrichment Designs for Two Subpopulations

Comparing Two Treatments vs. Control

Josh Betz, Jon Steingrimsson, Adi Gherman, & Michael Rosenblum   Johns Hopkins Bloomberg School of Public Health

June 13th, 2017

Multi-Arm Adaptive Enrichment Designs

- Suspected treatment effect heterogeneity: e.g. Molecular targets in Cancer
  - Treatment Effect varies across subgroups in population
- Enroll broadly initially, modify in pre-planned manner based on accrued data
  - Pre-specified Subgroups: defined prior to randomization
  - Efficacy/Futility at Interim Analysis: Group Sequential Methods
- Ethical & Efficient: common control; interim stopping
- Potential to reduce time to market, improve patent life
- Not guaranteed to be better than standard design

Optimization of Adaptive Enrichment Designs

- Most prior research on two-arm trials
- Exception: Wason and Jaki (2012)
  - 6 parameterized designs, 3 treatment scenarios; Binding futility
- Our Software: Generalized Optimizer for Enrichment Designs
  - Two treatments vs. Control
  - Familywise Type I Error Rate Control by design; Non-binding futility stopping
  - Compare performance across user-specified treatment scenarios
  - Continuous, Binary, and Survival Outcomes
- Application to Cardiac Resynchronization Therapy trials

Design Optimizer Workflow

- User Interface via Web
  - Optimization goal, clinically meaningful treatment effect (MCID), Subpopulation sizes, accrual rate, delay, power type
  - Each Scenario: outcome parameters, power constraints, weight
- Design modules translate user input into design parameters
  - Designs that strongly control FWER - Step-Down Dunnett: α-allocation, stages, futility boundaries, interim timing
  - Wrapper maps design parameters to performance
- Optimizer searches design space subject to user constraints
  - Modular: can optimize any design module + wrapper
- Re-evaluate performance upon convergence
- Reproducible report: design parameters & performance

Example Design Module: Step-Down Dunnett

- At each interim stage in each subpopulation:
  - Test non-stopped treatments for efficacy, then futility
  - If arm is stopped, patients simply are not enrolled.
  - Continue enrollment if at least one treatment remains
- Efficacy boundaries are similar to Spiessens and Debois (2010).
- Futility stopping is non-binding (Liu and Anderson 2008)
- Efficiency: leveraging covariance due to common control

Design Optimization

- Single Parameter Design: Binary Search
  - One stage, equal α allocation: Feasible sample size
- Simulated Annealing (SA) - Multiparameter Designs
  - 5000-7500 iterations: dimensionality
  - 10,000 simulations for design characteristics per iteration
- Starting Value of SA
  - 130% of Feasible One-Stage Equal-α Sample Size
  - Equal α-allocation
  - Futility Boundaries: $Z = -4$
  - Interim Analysis at 50% of Maximum Sample Size
- Distributed across 10 computing nodes; Seeded for replicability

Designs Implemented:

- Single Stage Equal α design: Sample size
- Single Stage Optimal α design:
  - 3 Parameters: Sample size, α-allocation (2)
- Two Stage Equal α design
  - 5 Parameters: Sample Size, Futility boundaries (4)
  - Interim Analysis at 50% Information Time
- Two Stage Optimal α
  - 10 Parameters: Sample Size, Futility boundaries (4), α-allocation (4), interim analysis time

Application | SMART-AV Trial: Cardiac Resynchronization Therapy

SMART-AV Trial - Rationale (Stein et al. 2010)

- Cardiac Resynchronization Therapy + Defibrillator (CRT-D): Patients with medically-refractive heart failure (HF) with severe left ventricular systolic dysfunction (LVESD).
- Timing of atrioventricular (AV) delay may improve disease progression, survival, hospitalization risk, HF symptoms, quality of life
- SMART-AV: Multi-center RCT Evaluating:
  - Doppler Echo-guided optimization (DEO)
  - SmartDelay algorithmic optimization (SDO)
  - Fixed Delay - No optimization (Control, Standard of Care)
- Suspect treatment effect heterogeneity based on disease severity
- Short QRS ($\leq$ 150 ms - healthier, greater chance to benefit) vs. Long QRS ($>$150 ms - More severe HF)

SMART-AV Trial - Design: (Stein et al. 2010)

- Objective: Evaluate within-subject 6 month change in LVESV
- Minimum Clinically Important Difference: 15 mL decrease in Left Ventricular End Systolic Volume (LVESV)
- Power Calculation based on two-sample t-test between optimized and fixed delay groups: $\sigma = 60$ mL: N=759
  - Assumed no difference between DEO & SDO
  - 'Internal pilot' with blinded sample size reassessment at N=75 assess variability in outcome; No interim efficacy analysis.
- Primary Analysis: ANCOVA: change in LVESV adjusted for baseline LVESV.
  - Superiority: SDO vs. Fixed; DEO vs. Fixed;
  - If SDO is superior to Fixed, assess non-inferiority of SDO to DEO;
  - If SDO is non-inferior to DEO, assess superiority of SDO to DEO

Adaptive Enrichment Design

- 2 Treatments (Optimization by Doppler Echo or SmartDelay) vs. Control (No optimization):
- 2 subpopulations: Short QRS ($\leq$ 150 ms) vs. Long QRS ($>$150 ms)
- Up to 2 Stages: Interim & Final Analysis
- For each treatment t and subpopulation s,
  - δ_{st} denotes effect of treatment a in subpopulation s
  - $H_{st} : \delta_{st} \leq 0$ for each $s \in \{1, 2\}, t \in \{A, B\}$ with strong control of FWER
  - Power $\geq$ 100(1-β)% to reject H_{sa} when $\delta_{st} \geq \delta_{min}$
  - Enrollment modification rule: if $f_{st} < Z_{st} < e_{st}$ at the end of stage 1, continue accrual in stage 2 for arm a and control in subpopulation s; Otherwise stop for efficacy/futility; Non-binding stopping for futility
- Minimize expected sample size under power constraints and compare operating characteristics of designs

SMART-AV Trial: (Stein et al. 2010)

- Short QRS: S_1 (49%); Long QRS: S_2 (51%);
- Primary: 6 Month LVESV Change (mL) - Continuous
- Secondary: NYHA Functional Class Improvement - Binary
- Delay = 6 months; Accrue 20 patients/month
- Fixed sample size vs. Two stage; Equal α allocation vs. optimized allocation;
- Size = 0.05; Power=0.8 All Non-Null

Simulation Scenarios

Scenario	δ_{1A}	δ_{1B}	δ_{2A}	δ_{2B}
1. Neither treatment effective - Global Null	0	0	0	0
2. A effective in s_1	δ_{min}	0	0	0
3. A, B effective in s_1	δ_{min}	δ_{min}	0	0
4. A effective in s_1, s_2	δ_{min}	0	δ_{min}	0
5. A effective in s_1, s_2; B effective in s_1	δ_{min}	δ_{min}	δ_{min}	0
6. A, B effective in s_1, s_2	δ_{min}	δ_{min}	δ_{min}	δ_{min}

- Asymmetric - A or B effective in s_2 if effective in s_1

Continuous Outcome - LVESV

Means

Scenario	Weight	C1	C2	A1	A2	B1	B2
1	0.167	0	0	0	0	0	0
2	0.167	0	0	15	0	0	0
3	0.167	0	0	15	15	0	0
4	0.167	0	0	15	0	15	0
5	0.167	0	0	15	15	15	0
6	0.167	0	0	15	15	15	15

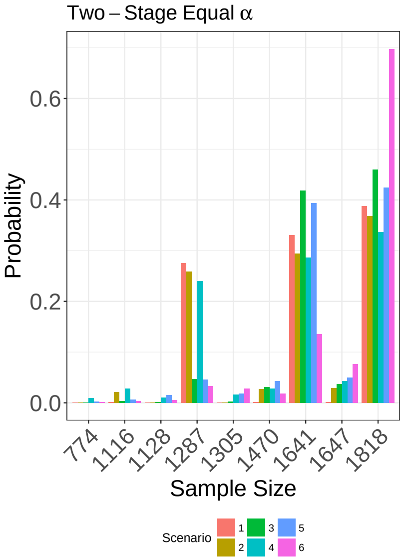
SDs

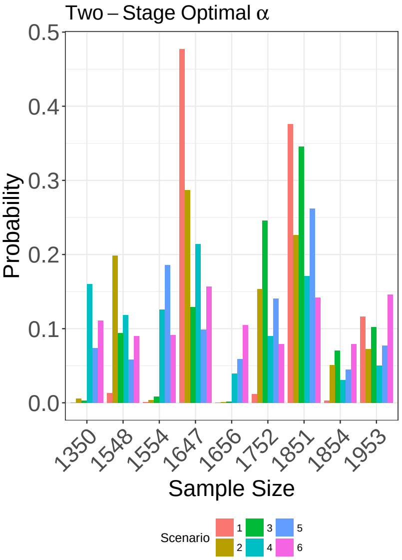
Scenario	Weight	C1	C2	A1	A2	B1	B2
1	0.167	60	60	60	60	60	60
2	0.167	60	60	60	60	60	60
3	0.167	60	60	60	60	60	60
4	0.167	60	60	60	60	60	60
5	0.167	60	60	60	60	60	60

Results - LVESV: Continuous Outcome

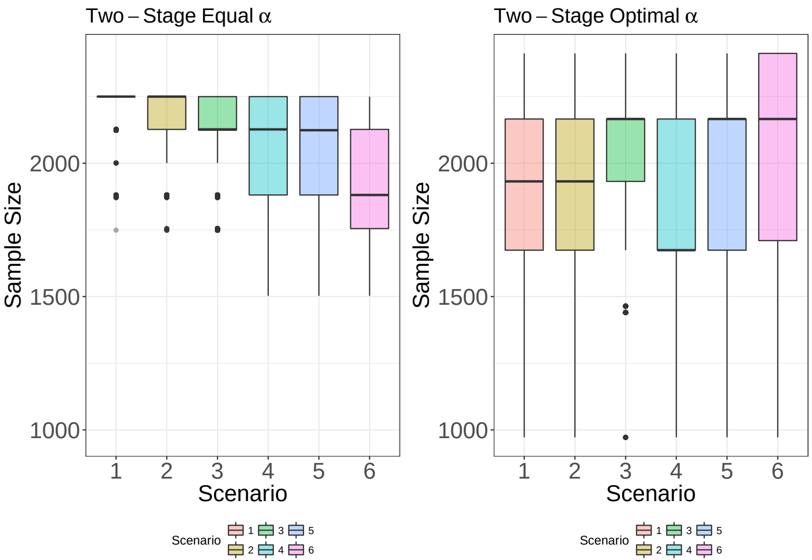
- One Stage Equal α: N=1827
- One Stage Optimal α: N=1782
  - $\alpha_{1(1)}$= 0.55; $\alpha_{2(1)}$= 0.45;
- Two Stage Equal α:
  - ESS=1716.7; MSS=1953
  - f_{1A}= -6.8; f_{1B}= -2.13; f_{2A}= 0.95; f_{2B}= 0.07;
- Two Stage Optimal α:
  - ESS=1648.6; MSS=1818; Interim: 0.22%
  - $\alpha_{1(1)}$= 0.09; $\alpha_{1(2)}$= 0.47; $\alpha_{2(1)}$= 0.05; $\alpha_{2(2)}$= 0.4;
  - $e_{1(1)}$= 2.85; $e_{1(2)}$= 2.22; $e_{2(1)}$= 3.02; $e_{2(2)}$= 2.29;
  - f_{1A}= -3.06; f_{1B}= -5.56; f_{2A}= -0.15; f_{2B}= -0.15;

Results - LVESV: Continuous Outcome





Results - LVESV: Continuous Outcome



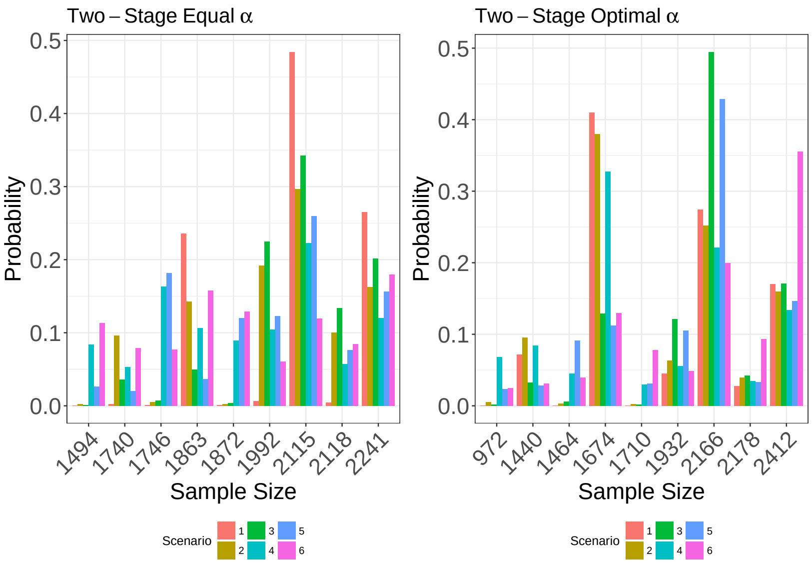
Binary Outcome - NYHA Class Improvement

Scenario	Weight	C1	C2	A1	A2	B1	B2
1	0.167	0.7	0.7	0.7	0.7	0.7	0.7
2	0.167	0.7	0.7	0.8	0.7	0.7	0.7
3	0.167	0.7	0.7	0.8	0.8	0.7	0.7
4	0.167	0.7	0.7	0.8	0.7	0.8	0.7
5	0.167	0.7	0.7	0.8	0.8	0.8	0.7
6	0.167	0.7	0.7	0.8	0.8	0.8	0.8

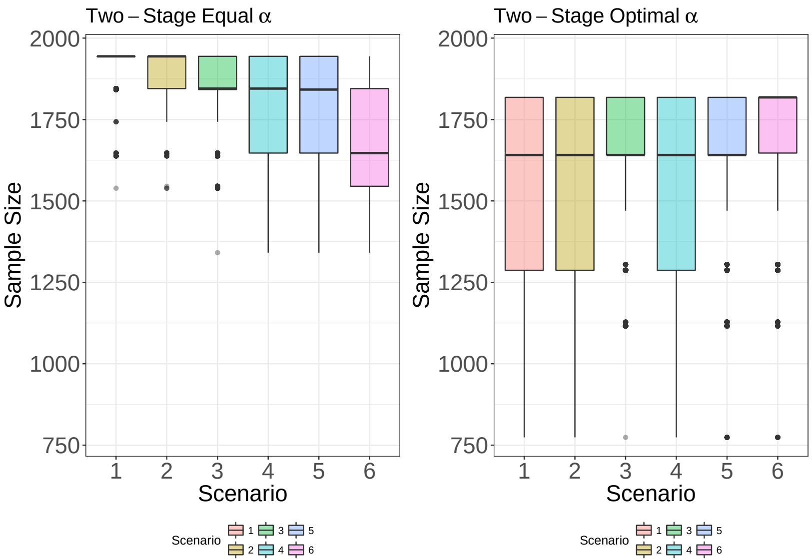
Results - Binary Outcome

- One Stage Equal α: N=2115
- One Stage Optimal α: N=2052
  - $\alpha_{1(1)}$= 0.53; $\alpha_{2(1)}$= 0.47;
- Two Stage Equal α:
  - ESS=2010.7; MSS=2241
  - f_{1A}= -6; f_{1B}= -2.86; f_{2A}= 0.5; f_{2B}= -0.62;
- Two Stage Optimal α:
  - ESS=1975.4; MSS=2412; Interim: 0.25%
  - $\alpha_{1(1)}$= 0.26; $\alpha_{1(2)}$= 0.24; $\alpha_{2(1)}$= 0.26; $\alpha_{2(2)}$= 0.25;
  - $e_{1(1)}$= 2.47; $e_{1(2)}$= 2.45; $e_{2(1)}$= 2.46; $e_{2(2)}$= 2.43;
  - f_{1A}= -1.11; f_{1B}= -7.12; f_{2A}= 0.31; f_{2B}= 0.38;

Results - Binary Outcome



Results - LVESV: Continuous Outcome



Future Directions

- Implementing additional designs
- Improved optimization techniques; Optimizing SA

Acknowledgements

- Funded by Patient-Centered Outcomes Research Institute (ME-1306-03198) and the US Food and Drug Administration (HHSF223201400113C).
- We would like to thank Boston Scientific for access to SMART-AV data
- We would like to thank Estelle Russek-Cohen, Shaila Shaheed, York Tomita, Elizabeth Colantuoni, Marina Kondratovich, Greg Levin, Amal Manseur, Sean Wybenga, and Jingyee Kou for their helpful comments.
- The views expressed in this talk are those of the speaker and not necessarily those of organizations above
- Conflicts of Interest: none

Optimization via Simulated Annealing

- Optimizer searches for optimal design over large parameter space: sample size, α-allocation, time of interim analysis, futility boundaries
- Doesn't require differentiable objective function
- $Y^{(n)}$ Objective Function at iteration n
- Current parameters $X^{(n)}$ and 'Temperature' $t^{(n)}$
  - Generate new candidate design: $X^{(n+1)} \sim N\left(X^{(n)}, \left(t^{(n)}/t^{(0)}\right)^2\right)$
  - Compare to current design: Accept if $U(0,1) < e^{\left(fracY^{(n+1)} - Y^{(n)} t^{(n)}\right)}$
  - 'Cool' system after a fixed number of candidates
- As system 'cools' search is more local and conservative

References

Liu, Q, and KM Anderson. 2008. “On Adaptive Extensions of Group Sequential Trials for Clinical Investigations.” *JASA* 103 (484): 1621–30.

Spiessens, B, and Debois M. 2010. “Adjusted Significance Levels for Subgroup Analyses in Clinical Trials.” *Contemp Clin Trials* 31 (6): 647–56.

Stein, KM, KA Ellenbogen, MR Gold, B Lemke, IF Lozano, S Mittal, FG Spinale, JE Van Eyk, AD Waggoner, and TE Meyer. 2010. “SmartDelay Determined AV Optimization: A Comparison of AV Delay Methods Used in Cardiac Resynchronization Therapy (SMART-AV): Rationale and Design.” *J Pacing Clin Electrophysiol* 33 (1): 54–63.

Wason, JMS, and T Jaki. 2012. “Optimal Design of Multi-Arm Multi-Stage Trials.” *Stat Med* 31 (30): 4269–79.